DOCUMENT RESUME

ED 038 812 48 EC 005 531

AUTHOR Singh, Surendra P.; Barnard, James
TITLE Evaluation of Instructional Materials - An Approach; A Position Paper.
INSTITUTION University of South Florida, Tampa.
PUB DATE Feb 69
NOTE 22p.

EDRS PRICE EDRS Price MF-$0.25 HC-$1.20
DESCRIPTORS *Evaluation, *Evaluation Techniques, Field Studies, Information Dissemination, *Instructional Materials, *Instructional Materials Centers, Models, Research Design, *Research Methodology

ABSTRACT

The role of the instructional materials center in evaluating materials is asserted and evaluation is defined as a media of compiling and analyzing information. Four evaluative methods are explained: field testing and comparing results; assessing rationale; measuring attitudinal and motivational variables; and describing the material and its use. A model for evaluation is presented; its six phases include decision making, identification and prescription, gathering information on predominant variables, posttests and compiling information, analysis (or evaluation) of information, and dissemination. Samples of models used in the field are appended. (JD)

ED038812

OE/BEH
BR 7-7
KC

Evaluation of Instructional Materials - An Approach

By

Surendra P. Singh, Ed.D.
Coordinator, Research and Evaluation Services, SMC
Assistant Professor
Special Education
University of South Florida

James Barnard, Ph.D.
Research Associate Professor
Institute III - Exceptional Children and Adults
University of South Florida

February 1969

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

EC005531F

Evaluation of Instructional Material - A Position Paper

During the last two years several special education instructional material centers have written position papers. In these papers (Ward, 1967, McIntyre, 1968, Olshin, 1968) primarily three issues have been discussed and described, i.e., the role of the center, the definition of evaluation and the evaluation procedure. All these issues are very complex and it does not seem easy to resolve them in one paper. Here, instead of taking issues with the controversies which prevail in the field, we would rather like to bring some of our views and doubts. We intend with every hope that this might help in clarifying our stand at the present time and generate some thought among individuals interested in the field.

As far as the first issue is concerned - the role of the center - we concur with W. C. Trow (1963) that the instructional materials center should promote learning rather than facilitate the administration's efficiency alone. Furthermore, we strongly believe that the center should act as an experimental distillary laboratory through which a well tested model for evaluation can be produced and kept updated.

Second issue - the definition of evaluation - is not in any respect less complex. Since our goal is not to analyze the ramification of different definitions, we would like to concur with

Cronbach's definition with an addition for our purpose. Cronbach (1903) had conceived evaluation as "the collection and use of information to make decision about an educational program". Cronbach's definition has emphasized the collection of information. While we agree with his contention of compiling information, we also feel that collection of information isn't enough. Therefore we would propose an extended definition of evaluation. It seems it would be more sensible not only to involve ourselves in collecting the information but also in sharing the responsibility of analyzing the information. Briefly, we perceive evaluation as a media of compiling and analyzing information and the Southeastern Materials Center as an instrument through which this can be schematized and processed to promote better learning.

The third issue - the evaluation procedure - although would encompass the first two as well, perhaps needs more attention. The most common way in which a material is evaluated is to gather date on its efficacy before and after it has been used in a field testing situation. The educational material, let's say a reading program, produces a certain measurable result and this is compared with the results produced by other reading programs or with the results produced in a control situation where no program at all had been used.

A second way in which a material may be evaluated is in terms

of its rationale, how clearly it is described from the rationale, and how much sense the rationale makes based on the state of our knowledge about reading at the moment of evaluation. Third, a material can be evaluated in terms of the attitudes of both the teacher-user and the student consumer have toward it, and fourth, evaluation can refer to the simple description of the materials and to an actual account of their use. This paper will discuss some of the ramifications of each of the above four ways in which the evaluative process may be approached.

In the empirical approach to evaluation, some attempt is made to gather data that will bear upon the question of the efficacy of the material in question. This data represents students' performance levels at some specified time after having been exposed to the material in some form. The data is then used in a comparison with other data representing performance levels of students who have not had the benefit of the material in question. There are at least three ways in which this comparison can be made. First, the comparison can be made between those that have been exposed to the program and those that have been exposed to no program. If the student progresses to a certain point within the program in a specified amount of time, this point can be compared with the progress of students who have not been introduced to a systematic reading program. Second, a comparison can be made

between performance levels of students having the target reading program and students having been subjected to other systematic reading programs. Although we know of no research relevant to the efficacy of the SRA reading program, nevertheless, it is highly likely that the above first comparison has already been made. That is, one could easily demonstrate that a reading program such as the SRA program is better than no program at all. The problems of this particular kind of comparison resides in the fact that it really isn't practically possible over a long period of time to have a control group that is constituted of children who are given no systematic training in reading. This simply would be unethical. This problem can be researched over relatively short periods of time however, and undoubtedly has already been done using within group design.

Apparently much of the research relevant to the second type of comparison, that among two or more specific reading programs, has shown limited success in demonstrating the efficacy of one reading program over another. Studies such as the one done by Woodcock at Peabody in which he compared five approaches to reading with primary level EMR's and found no differences among the groups supports the above rather pessimistic notion that individual differences in reading are not due to the relative efficacy of one reading program over another.

The above second comparison, to be made among reading programs, leaves the evaluator with a very difficult set of problems. The above problem of small differences in performance level being accounted for by the specific nature of the reading program used, is only one aspect of the difficulty. Another problem is the fact that when working with mentally retarded children, one must expect that they have already been exposed to a number of attempts to train them to read. This means, that since they may well still be virtual non-readers, these previous attempts at teaching them to read were unsuccessful in improving their ability to handle themselves well in typical reading situations, though of course, they may have shown some ability to handle the material specific to the reading program. A further problem is that our assessment instruments may not be fine grained enough to identify whatever small differences among programs that may exist.

Another problem exists in this type of between group comparison. It is very likely that two different reading programs will have different goals and will certainly proceed along different avenues to reach their different goals. This would mean that the comparison between the two programs would be meaningless, at least to the extent that the two programs taught different skills. One might argue that the goal of all programs of reading is to teach a certain general proficiency when one encounters typical reading

situations. Thus, all programs would involve stress on vocabulary, at least to some extent, and so for those children who were able to go through the entire reading program, they would find themselves able to handle these typical reading situations. It is doubtful whether we could expect experts to agree to what level of reading proficiency that all reading programs should lead. Even if they could agree, however, there is still the problem that when we deal with mentally retarded children we must expect there to be a great range in the proficiency of their reading skills. This means that many of these children will not get through the entire reading program. Even if we had programs of differing difficulty and used them with the appropriate children, there would still be children that could not get through their reading program. This means that one would be left in the position of comparing reading proficiency among groups of children that had not completed their programs and were at various stages of the program. Thus, a particular child's performance would be directly related to the individual reading program in which he had been trained. For instance, a child that had gone through half of the SRA reading program would not be familiar with some very common sight words, such as pronouns, as these words do not occur until much later in the program. However, other reading programs include pronouns very early in the lessons.

The above seems to point to the conclusion that the intermediate goals set up by reading programs are not directly comparable across programs and in that case, obviously, student performance can not be meaningfully compared across programs.

So far the discussion has dealt with between group comparisons. A third procedure involves comparisons within a single group. Comparisons are made between students' initial and final performances, with the manipulation, in this case the reading program, sandwiched in between. The question asked is: Does the student increase reading efficiency from the point at which he started as a function of having been trained with a particular reading program?

In each of the above three comparison situations, the experimental vs. control groups comparison, the between experimental groups comparison, and the within group subject as his own control comparisons, there are at least three methods of documenting performance increase. The first method has to do with what might be called the distance travelled. That is, simply, how much progress has an individual shown within a certain content area as a function of having been introduced to the reading materials in question. The specific tests used to assess distance travelled are carefully chosen to reflect the specific material contained in the reading program, thus they are program specific. The

second method of documenting performance increase has to do with the retaining of what was learned over a specified amount of time. This can be assessed over short periods of time such as would exist between successive lessons, or over longer periods of time such as would exist over a summer vacation. Typically, retention is measured either in terms of recall or of relearning criteria.

Finally, the third method of ascertaining performance increase is in terms of the degree of transfer that is attendant upon the use of the particular material in question. This criterion has been discussed above and has to do with the degree of success an individual experiences after he has completed an entire reading program in what can be termed typical reading situations. Obviously, no individual has been trained specifically to handle all possible reading situations and therefore it is expected that he will be able to transfer his ability to perform successfully in the programmed reading situations to new reading situations.

Going back to three ways in which comparisons can be made, two between-group comparisons and one within group comparison, it would appear that the latter may turn out to be the most useful, as it gets around a number of difficulties as above outlined. However, it has one difficulty of its own that needs to be discussed. This is the problem of relativity. With this within group method of comparison, it is possible to document change across time as a function of having encountered a specific reading

program, but is this change adequate? We have no readily available guidelines that tell us when this change is enough. This brings us to a discussion of the second major mode of evaluation of educational materials. This mode is a rational approach and simply has to do with the rationale upon which the particular materials in question are based. The most meaningful comparison that can be made is made not between reading programs based upon different rationales, but rather upon the actual results obtained with individuals subjected to the particular reading program in question within the context of the expectations as laid down in the rationale upon which the program is based. This type of evaluation must of necessity be made by experts in the field from which the program has been taken as they would be in the best position to judge the goodness of a particular rationale. These judgmental criteria might be based on such questions as: Does the rationale fit with the current research evidence? Is the rationale fairly complete and are the various component steps clearly and logically elaborated and interrelated?

A third major mode in which the evaluation of materials' questions may be asked has to do with attitudinal and motivational variables of both the teacher and the students. Based upon an acquaintance of the research having to do with the specific area of the teaching of reading, it can be hypothesized that the degree of enthusiasm that the teacher has for the particular approach to

reading that he is using is one of the major determinants of success or failure in that teaching. That is, if he is convinced that a particular approach is a good one, it is more likely that he will have success with it than if he has basic doubts abouts its appropriateness. This would suggest that one way in which one might increase the efficacy of this type of material would be to change the attitude of the user in the direction of the users being more convinced in fact that it is a good program. Very likely, there is a good deal of research already done that would be very relevant to this question particularly in the area of attitude change. There are things that can be done to change the attitude of individuals concerning various situations important to them. The basic question here is, of course, if you have an individual who is positively disposed towards the material, can the material be used more effectively than if the individual using the material is more negatively disposed to it. If this in fact can be clearly demonstrated, then the next question is how do you manipulate an instructor's disposition towards the material? One could study the second question by essentially manipulating the packaging of the materials sent to various potential users. The relevant dimensions of packaging are of course themselves open to research but would probably include such things as physical aspects of packaging, the attractiveness of the material, and also the types of instructions and written explanations that

would accompany the material. One could design a study where materials would be sent to carefully selected teachers over the state which were accompanied by materials of packaging. After the materials had been sent, and the teacher had had a chance to look them over, an assessment instrument could be introduced which would be designed to look at the instructor's disposition toward the materials. This measure would then be related to students' gains in the area relevant to the materials being tested. Ways in which one might manipulate the packaging of materials could include the following: First, statements of endorsement by well known individuals and second, it might be possible to present a distillation of research evidence that demonstrates in some way that in fact the program works. There would, of course, be several ways in which the motivation of the teacher could be manipulated within the context of the use of the materials themselves. It would seem, for instance, that it would be important for a teacher to receive very frequent reinforcements for the use of the program in the form of student progress. Thus, a teacher would feel better disposed towards a program if he could observe progress after each small unit of work rather than waiting for confirmation of success until the end of the day.

Assessment of the attitudes of the students engaged in the reading programs also would obviously be very important in the determination of the success of that program.

The fourth mode in which the evaluation question can be approached has to do with a description of the use of the materials by the teacher in his day to day encounter with his students. In discussions with teachers, it has become obvious that they are very concerned with exactly how they will be able to handle these materials within their specific classrooms.

Following is a list of specific considerations that might be used in a questionnaire designed to tap this fourth mode of evaluation:

1. Who developed the material in question, and where is it available?
2. What does the material cost?
3. For what ages is it used most appropriately?
4. For what student characteristics is this material best used?
5. Does this material have an intrinsic interest value?
6. What are the physical characteristics of the material (in terms of quality and durability)?
7. What specific content areas are involved in the material?
8. What are the characteristics of the teacher who used the material (training, etc.)?
9. How large was the size of the class in which the material was used?
10. Are the instructions that accompany the materials sufficiently clear and complete?
11. Is it possible to integrate the material in question with other materials used during the day?

12. Is the rationale upon which the material is based clearly given and is it educationally sound?

13. What teaching objectives does the material meet, (to contrast, to describe, etc.)?

14. Can the materials be used with groups of children? How much individual time is required in the use of the materials?

15. Is the material rigidly sequenced or does it allow of several novel variations?

16. Is it difficult to adjust rate of progress to meet individual students' needs?

17. How does this material compare with other similar materials you have used?

18. What specific difficulties did you experience when you used the material?

It is quite possible that these questionnaires could be made available to prospective users as well as to the SMC staff.

In summary, there are at least four major areas in which evaluation may proceed. First, the evaluation can be made in terms of the efficacy of a particular set of materials. This phase of the evaluation is based upon empirical evidence gained from teachers actually using materials with relevant individuals. Second, evaluation can be made in terms of the rationale upon which a particular set of materials are based. Third, an evaluation procedure can include an assessment of motivational and attitudinal variables relevant to the use of the materials. And fourth, an evaluation can include a description of the nitty-gritty encountered in the front lines by the individuals who are actually using the materials.

A Proposed Model for Evaluation

The above discussion provides us with a picture of complexity of variables which surmounts the field of evaluation of instructional materials. This further demands to build a schematic model best suitable to deal with the problem at hand.

There are several models such as developed by EPIE, Consumer's Union, etc. (See Appendix A) which are presently being used. One of them although, has great merits but seems complex and time consuming. The second one wasn't built for educational purposes in the first place. Here we are proposing an evaluation diagram with six phases (See Diagram No. 1). This we hope will facilitate us to reach our goals:

> "...1) to develop and evaluate instructional materials at the center, and
> 2) to procure and disseminate the information on widely used and needed materials to make the evaluation procedure more meaningful and relevant."*

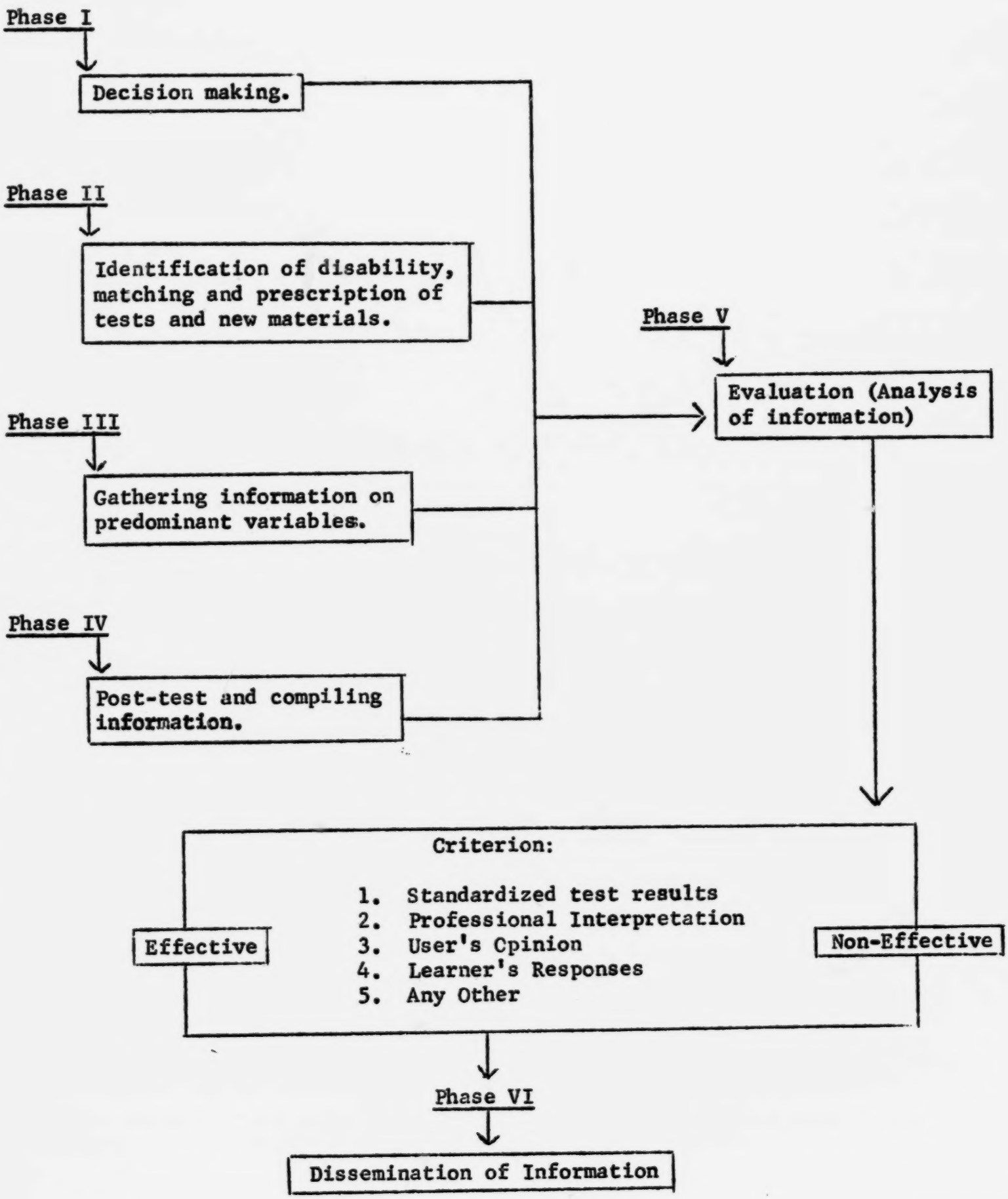
Phase I - Decision Making Phase: The research and evaluation services unit within the instructional materials center will have to make decisions in regard to what to evaluate. Different centers might suggest different approaches to resolve this question. It wouldn't be out of place to describe how we tried to face this important question.

At the time of onset of research and evaluation unit in September 1968, we had two major questions to resolve. No. 1 - In what area we shall concentrate? No. 2 - What criterion shall we set up for

---

*Singh, S.P., SMC Service is Instrument for Research and Evaluation. SMC News Notes, Vol. II, No. 5, December, 1968.

(SMC) EVALUATION DIAGRAM NUMBER 1



the priority? Since we strongly believe in complete involvement of users and the individuals in the field, a pre-planning session meeting of Special Education County Coordinators in the state of Florida was initiated at the Southeastern Materials Center on November 7 and 8, 1968. During this meeting, it was felt that there is a pressing need to evaluate the vast numbers and types of materials in the area of language development and perceptual training. Since we believe that the meeting was well attended and represented, the view of the users in the field (See Appendix B), we could very well devote ourselves to these areas of need and needless to say, this covers alot of field and would undoubtedly keep us busy.

For the second question of setting up priority, we have set up the following criterion. This is based on the SMC's goals and policies discussed at the staff meetings and experience gained at the evaluation planning meeting.

1. Materials* developed at the Center.	First Priority
2. Materials advised by the Advisory Committee. (See Appendix C)	Second Priority
3. Materials suggested by the individuals in the field within the SMC region.	Third Priority
4. Materials suggested by other network centers.	Fourth Priority
5. Materials developed at SMC or suggested by the Evaluation Advisory Committee, individuals in the field within the region and network centers not related to language development or perceptual training.	Fifth Priority

Phase II - Identification and Prescription: Once decided what to evaluate, every effort will be made to match the instructional material and/or program under study to appropriate disability group.

---

*Materials related to language development and perceptual training.

The decision for appropriateness matching will be made on the basis of the author's and publisher's report, plus any other report including research which might make matching more objective. During this phase, besides matching instructional materials to the disability group, appropriate pre-tests and SMC questionnaires and forms (See Appendix D) will also be ascertained and prescribed.

Phase III - Gathering Information on Predominant Variables: There are at least three important elements besides instructional materials which would need consideration to make any decision in regard to the effectiveness of materials on pupils. The first element is the teacher who will be using the materials in the class. It seems to be of great importance to know about a person who will be using the material. Any information such as his experience, formal education, personality, self-concept, attitude and interpersonal relationship with faculty, administration and students would be worthwhile. The second element is the school, the place where the material is being used. The information such as the educational philosophy of the school, the facilities available, the size, etc. should also be gathered. The third element in which we would be interested is the community. The knowledge about the size of the community, the location, the type, the predominant age group, the educational level, the political affiliation and the attitude toward the education will facilitate the interpretation of the instructional material used in the school.

Phase IV - Post-tests and Compiling of Information: During this phase, the information will be collected on the standardized tests, questionnaires, learner's response and any other instrument prescribed during the second phase.

Phase V - Analysis of Information: Analysis of information for evaluation will be made at the Southeastern Materials Center. The analysis will utilize all the information gathered on tests and other instruments. The nature of analysis will be determined by the professional staff of SMC and the consultants according to the individual projects.

Phase VI - Dissemination of Information: This is an important phase. It should be used to pump back the information to the users and the individuals in the field. This could be achieved through the SMC News Notes, professional journals and publishing yearly SMC research monographs.

In summary, although the entire field of research and evaluation is complex and difficult, there are certain ways to simplify and make it effective. We feel the six phase approach would facilitate the evaluation of instructional materials.

REFERENCE

1. Denny, T. Educational Product Evaluation. First Annual Project ARISTOTLE Proceedings, 1967, in Press.

2. McIntyre, R.B. and Nelson, C.C. The Evaluation of Instructional Materials in a Systems Framework. Paper submitted to the Evaluation Conference, April 4, 5, 6, 1968, SEIMC George Washington University, Washington, D.C.

3. Stufflebeam, D.L. A Depth Study of the Evaluation Requirement. Theory into Practice, 1966, 5, 121-133.

4. Trow, W.C. Teacher and Technology. New York: Meredith, 1963, 198pp.

5. Van Dalen, D.B. Understanding Educational Research. New York: McGraw Hill, 1962.

6. Ward, T. and Levine, J. Stimulation and Recording Device for Research in Programmed Instruction - The Prospectus Series No.5-Michigan - Indiana - Ohio: USOE/MSU Regional Instructional Materials Center for Handicapped Children and Youth, 1967.

7. Wittrock, M.C. Paradigms in Research on Instruction. Educational Psychologist, 1967, (1), 1-3.

8. _____. Proceedings on Conference for the Evaluation of Instructional Materials. Washington: SEIMC, The George Washington University, 1968.

APPENDIX A

Samples of Models Used in the Field

F2005 331E

CONSUMER UNION'S MODEL*

EVALUATION OF PRODUCTS

DECIDING PROCEDURE

WHAT PRODUCTS ARE TO BE EVALUATED

I) Operation committee
- a) Marketing group
- b) Library and information group
- c) Technical department
- d) editorial department

WHICH BRAND MODELS TO EVALUATE

II)
- a) Product most widely available
- b) Warranty practices
- c) Price information
- d) Delivery and installation problems, etc.

WHAT TO TEST AND HOW TO TEST

III)
- a) Library information, the standards and specifications that have been developed for such products all over the world.
- b) Chemists and technologists write to sources for information, criteria, & methods.
- c) Sophisticated tester will devise the test procedures.

WHAT CHARACTERISTICS TO LOOK FOR

IV)
- a) Performance
- b) Convenience
- c) Safety
- d) Durability
- e) Cost

ITEMS LEFT FOR CONSUMERS CHOICE

V)
- a) Style
- b) Appearance, etc.

EVALUATION PROCESS

I) Laboratory Testing

II) Small "Panel" use test

III) Field Trial

IV) Dissemination of data
- a) Nobody wants a product that is unsafe.
- b) Nobody wants a product that will fall apart after a week of use

V) Editorial Board (report writing)

---

*Adapted from article by Kaplan, M. The Consumers Union Model. Proceedings of The Conference for the Evaluation of Instructional Materials, Washington, D.C., April 5-6, 1968.